

Review

# Predicting Spinal Cord Injury Prognosis Using Machine Learning: Systematic Review and Meta-Analysis

Linxing Zhong, MS; Qiying Huang, MS; Hao Zhang, MD; Liang Xue, MS; Yehuang Chen, MS; Jianwu Wu, MS; Liangfeng Wei, MD

Department of Neurosurgery, Fuzong Clinical Medical College of Fujian Medical University, Fuzhou, China

**Corresponding Author:**

Liangfeng Wei, MD  
Department of Neurosurgery  
Fuzong Clinical Medical College of Fujian Medical University  
156 West Second Ring North Road  
Fuzhou 350025  
China  
Phone: 86 13960760177  
Fax: 86 591-87640785  
Email: [Wlf74@163.com](mailto:Wlf74@163.com)

## Abstract

**Background:** Spinal cord injury (SCI) is complicated and varied conditions that receive a lot of attention. However, the prognosis of patients with SCI is increasingly being predicted using machine learning (ML) techniques.

**Objective:** This study aims to evaluate the efficacy and caliber of ML models in forecasting the consequences of SCI.

**Methods:** Literature searches were conducted in PubMed, Web of Science, Embase, PROSPERO, Scopus, Cochrane Library, China National Knowledge Infrastructure, China Biomedical Literature Service System, and Wanfang databases. Meta-analysis of the area under the receiver operating characteristic curve of ML models was performed to comprehensively evaluate their performance.

**Results:** A total of 1254 articles were retrieved, and 13 eligible studies were included. Predictive outcomes included spinal cord function prognosis, postoperative complications, independent living ability, and walking ability. For spinal cord function prognosis, the area under the curve (AUC) of the random forest algorithm was 0.832, the AUC of the logistic regression algorithm was 0.813 (95% CI 0.805-0.883), the AUC of the decision tree algorithm was 0.747 (95% CI 0.677-0.802), and the AUC of the XGBoost (extreme gradient boosting) algorithm was 0.867. For postoperative complications, the AUC of the random forest algorithm was 0.627 (95% CI 0.441-0.812), the AUC of the logistic regression algorithm was 0.747 (95% CI 0.597-0.896), and the AUC of the decision tree algorithm was 0.688. For independent living ability, the AUC of the classification and regression tree model was 0.813. For walking ability, the model based on the vector machine algorithm was the most effective, with an AUC of 0.780.

**Conclusions:** The ML models predict SCI outcomes with relative accuracy, particularly in spinal cord function prognosis. They are expected to become important tools for clinicians in assessing the prognosis of patients with SCI, with the XGBoost algorithm showing the best performance. Prediction models should continue to advance as large data are used and ML algorithms develop.

*JMIR AI*2025;4:e66233; doi: [10.2196/66233](https://doi.org/10.2196/66233)

**Keywords:** artificial intelligence in health care; spinal cord injury; predictive analytics; clinical outcomes research; PRISMA; Preferred Reporting Items for Systematic Reviews and Meta-Analyses

## Introduction

Spinal cord injury (SCI) is one of the most devastating diseases. From 1990 to 2019, the global prevalence of SCI increased by 81.5% (74.2%-87.1%) and the incidence

increased by 52.7% (30.3%-69.8%), with an annual rise [1]. The injury has severe impacts on patients, potentially leading to varying degrees of motor, sensory, and autonomic dysfunction [2]. Furthermore, SCI imposes a heavy burden on the health care system and socio-economics. Statistics

show that the average lifetime rehabilitation cost for patients with SCI may exceed US \$750,000, resulting in an estimated annual expenditure of US \$6 billion for SCI in the United States [3]. Since there are still no effective treatments for SCI, the actual cost of treatment keeps going up every year. Due to the complexity of SCI, it has become a global public health issue, with significant public concern regarding its treatment, rehabilitation, and prognosis. The prognosis of SCI is especially important for clinical doctors. Early assessment of patients with SCI is crucial to prevent overtreatment and to provide early personalized treatment for those with favorable prognostic opportunities, thus improving patient outcomes and facilitating their return to home and society as much as possible. Prognosis assessment is a challenging undertaking for clinicians due to the complexity, diversity, and individuality of SCI.

Currently, clinical tools for SCI prognosis assessment largely rely on Abbreviated Injury Scale (AIS) scores [4] and Frankel grades [5], which involve evaluating the sensory and motor functions of patients after injury to roughly determine the nature of SCI and the patient's prognosis. However, the assessment process requires high patient cooperation, and the results can be easily influenced by factors, such as the timing of neurological examinations (eg, spinal cord concussion period), the patient's condition (eg, intoxication, sedation, and accompanying brain injury), and the subjectivity of the evaluator [6]. Machine learning (ML) is a significant area of artificial intelligence that has seen extensive application in clinical care in recent years due to the field's rapid progress. Park et al [7] used ML combined with low-dose CT to predict prognostic biomarkers and molecular subtypes associated with invasive breast cancer. Gupta et al [8] used ML combined with magnetic resonance imaging to detect and classify brain tumors and their stages. Jumper et al [9] used ML to construct AlphaFold for predicting protein structures. ML has demonstrated high accuracy and predictive ability [10], establishing more reliable predictive models through continuous integration and analysis of large amounts of complex, nonlinear data to assist clinical decision-making.

ML has attracted considerable attention from researchers regarding SCI prognosis. Nevertheless, there is a dearth of solid data supporting the efficacy of ML models in the prognosis of SCI, and there are no systematic reviews contrasting the variations and importance of various models and prognostic markers. Therefore, this study aims to fill this gap. We conducted a systematic review and meta-analysis to evaluate the performance and quality of ML models in predicting SCI prognosis.

## Methods

The systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Diagnostic Test Accuracy (PRISMA-DTA) guidelines. The study protocol has been registered in the PROSPERO

and was approved before the start of the study (ID: 42023481977).

## Search Strategy

We conducted a literature search for studies on ML-based SCI prognosis prediction published up to February 20, 2024. A comprehensive search was performed across the following 9 electronic databases: PubMed, Web of Science, Embase, PROSPERO, Scopus, China National Knowledge Infrastructure, Cochrane Library, China Biomedical Literature Service System, and Wanfang Data.

## Inclusion and Exclusion Criteria

Inclusion criteria were (1) study subjects are patients with SCI, regardless of the nature, degree, level, or time of SCI; (2) the research method involves ML, not limited to any specific algorithm or model; (3) the purpose of the study is to predict the prognosis of patients with SCI, including spinal cord function prognosis, complication occurrence, quality of life, etc; (4) the study provides performance metrics of the model, such as accuracy, sensitivity, specificity, area under the receiver operating characteristic curve, etc; and (5) published in Chinese or English. Exclusion criteria of the study were (1) concurrent cranial brain injury or injuries to other parts of the body outside the spinal column; (2) reviews, case reports, conference abstracts, expert opinions, or other non-original research; (3) studies that do not build ML models or do not provide relevant performance metrics for ML; and (4) articles for which full text or original data cannot be obtained, repeated publication, or data duplication.

## Data Extraction and Analysis

The literature retrieved from each database was imported into EndNote X9 (Clarivate) reference management software. Two researchers trained in the systematic review process performed the screening and data extraction, with cross-checking. Initially, articles were screened based on titles and abstracts, and then full texts were reviewed according to the inclusion and exclusion criteria. For information that was uncertain but crucial to the study, the original authors were contacted via email or phone. Following the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) guidelines for extracting data from predictive model studies [11], the following data were independently extracted from each study: study design, characteristics of the SCI population, types of ML models used, study outcomes, and predictive performance of various models.

## Risk of Bias Assessment

Two researchers used the risk of bias (ROB) assessment tool for predictive models (Prediction Model Risk of Bias Assessment Tool [PROBAST]) to evaluate the ROB of each model in every study and the applicability to the issues we reviewed [12]. Any disagreements were resolved by a third investigator. Each model was assessed in 4 domains (participants, predictors, outcomes, and analysis) as having "high-risk," "unclear," or "low-risk" based on a series of

specific questions. The same scale was used to evaluate the applicability of each model to the issues we reviewed in 3 domains (participants, predictors, and outcomes).

## Statistical Analysis

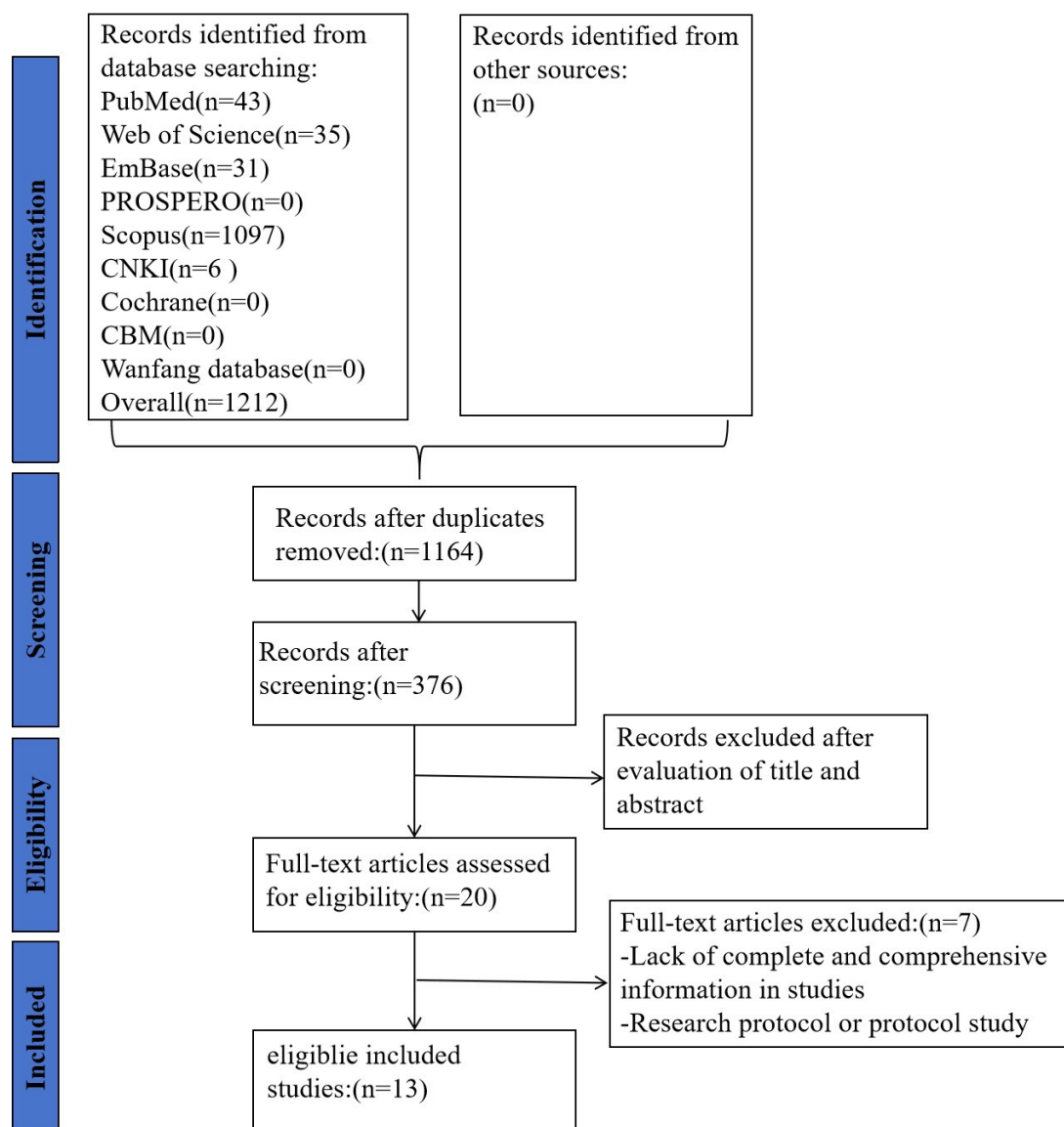
We used R software (version 4.2.3; R Core Team) to perform the meta-analysis. Based on the types and distributions of ML performance indicators, we selected appropriate effect metrics and models, such as 95% CI, and conducted a summary analysis using either a fixed-effects model or a random-effects model.

## Results

### Literature Selection Process

The research selection process was shown in [Figure 1](#). A total of 1212 unique records were identified, and after excluding studies without full text, conference abstracts, registration protocols, studies that failed to provide accuracy metrics for ML, and those that did not develop new ML models, 13 studies [13-25] were finally included.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for the research search process.



### Basic Characteristics of Included Studies

The 13 included studies were from Japan (6/13, 46.15%), the United States (2/13, 14.38%), India (1/13, 7.69%), Canada (2/13, 14.38%), South Korea (1/13, 7.69%), and China (1/13, 7.69%). The study types included retrospective cohort studies [13-15,17-24] (9/13) and prospective cohort studies (2/13) [15,25]. Most articles were published in 2023 (7/13) [13,15,16,18,21-23], followed by 2022 (2/13) [19,24], 2024 (1/13)

[20], 2020 (1/13) [17], 2019 (1/13) [25], and 2016 (1/13) [14]. A total of 51 ML models were, with 31 models for predicting spinal cord function prognosis, 15 models for predicting postoperative complications, 1 model for predicting independent living ability, and 4 models for assessing walking ability. The study characteristics and performance metrics are presented in [Table 1](#).

**Table 1.** Basic characteristics of included studies.

Study	Year	Area	Sample size	Design	Artificial intelligence model	Performance evaluation
Spinal cord function prognosis						
Maki et al [13]	2023	Japan	3122	Retrospective study	CatBoost, Gradient Boosting, RF <sup>a</sup> , Extra Trees, LR <sup>b</sup> , Ada Boost, Linear Discriminant Analysis, Light Gradient Boosting Machine, Extreme Gradient Boosting, Quadratic Discriminant Analysis, Naïve Bayes, K Neighbors, DT <sup>c</sup>	Accuracy, AUC <sup>d</sup> , recall, precision, $F_1$ -score
Belliveau et al [14]	2016	US	3142	Retrospective study	ANN <sup>e</sup> , LR	AUC, NLR <sup>f</sup> , PLR <sup>g</sup>
Kalyani et al [15]	2023	India	165	Retrospective study	XGBoost <sup>h</sup> , LR, DT	AUC, Accuracy
Facchinello et al [16]	2023	Canada	172	Prospective study	CART <sup>i</sup>	NA <sup>j</sup>
Inoue et al [17]	2020	Japan	165	Retrospective study	XGBoost, LR, DT	AUC, Accuracy
Shimizu et al [18]	2023	Japan	135	Retrospective study	Light GBM, XGBoost, CatBoost	Accuracy, AUC, recall, precision, $F_1$ -score
Okimatsu et al [19]	2022	Japan	215	Retrospective study	RF	Sensitivity, Specificity, Accuracy, $F_1$ -score
Kato et al [20]	2024	Japan	210	Retrospective study	RF, SVM <sup>m</sup> , NN <sup>s</sup> , XGBoost	None
Postoperative complications						
Li et al [21]	2023	China	870	Retrospective study	RF, XGBoost, GBM <sup>k</sup> , NBC <sup>l</sup> , DT	Sensitivity, Specificity, Accuracy, AUC
Luther et al [22]	2023	US	4709	Retrospective study	the gradient boosting, LR, the adaptive LASSO <sup>l</sup> model	AUC
Kim et al [23]	2023	Korea	623	Retrospective study	GNN-GCNP <sup>q</sup> , DNN <sup>o</sup> , SVM-linear, SVM_RBF <sup>r</sup> , KNN <sup>n</sup> , RF, LR	Sensitivity, Specificity, Accuracy, AUC, $F_1$ -score
Independent living ability						
Hori et al [24]	2022	Japan	1404	Retrospective study	CART	Sensitivity, Specificity, Accuracy, AUC, $F_1$ -score, the positive predictive value
Walking ability						
DeVries et al [25]	2019	Canada	862	Prospective study	LR, VM <sup>u</sup> , Hicks, Unsupervised MLA <sup>v</sup>	AUC, $F_1$ -score

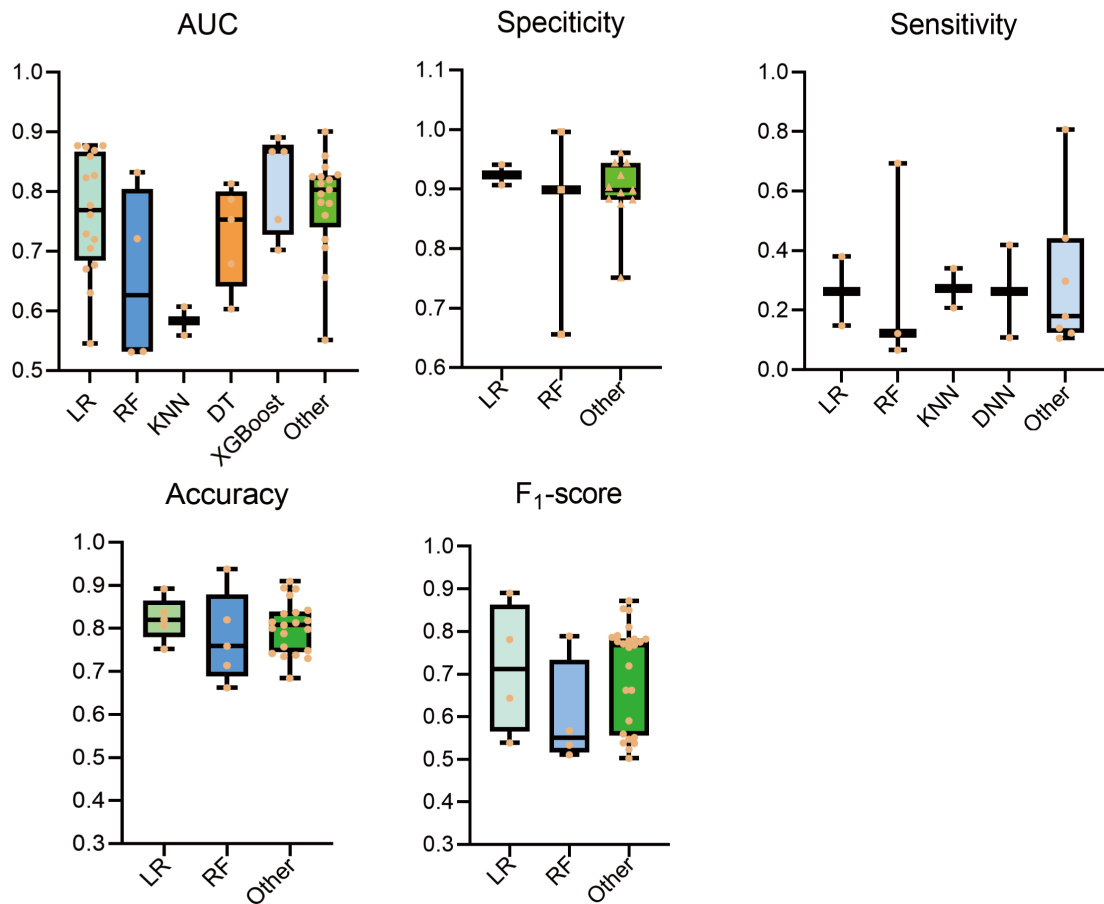
<sup>a</sup>RF: random forest.<sup>b</sup>LR: logistic regression.<sup>c</sup>DT: decision tree.<sup>d</sup>AUC: area under the curve.<sup>e</sup>ANN: artificial neural network.<sup>f</sup>NLR: neutrophil-to-lymphocyte ratio.<sup>g</sup>PLR: platelet-to-lymphocyte ratio.<sup>h</sup>XGBoost: extreme gradient boosting.<sup>i</sup>CART: classification and regression tree.<sup>j</sup>NA: \_\_\_\_.<sup>k</sup>GBM: gradient boosting machine.<sup>l</sup>NBC: naïve Bayes classifiers.<sup>m</sup>SVM: support vector machine.<sup>n</sup>KNN: K nearest neighbors.<sup>o</sup>DNN: deep neural networks.<sup>p</sup>GNN: graph neural networks.<sup>q</sup>GCN: graph convolutional network.<sup>r</sup>SVM\_RBF: support vector machine using a radial basis function.<sup>s</sup>NN: neural network.<sup>l</sup>LASSO: least absolute shrinkage and selection operator.<sup>u</sup>VM: vector machine.<sup>v</sup>MLA: machine learning algorithm.

## Evaluation Performance of Different Prediction Models

The area under the curve (AUC) has long been used to evaluate model performance and is therefore a primary

performance indicator. The AUC range for ML prediction models is 0.532-0.904 (Figure 2). Additionally, other performance metrics of ML prediction models, such as specificity, sensitivity, accuracy, and  $F_1$ -score, were also extracted.

**Figure 2.** Summary of performance for different machine learning models. AUC: area under the curve; DNN: deep neural networks; DT: decision tree; KNN: K nearest neighbors; LR: logistic regression; RF: random forest; XGBoost: extreme gradient boosting.



**Results of Bias Risk Assessment**

The ROB in ML prediction based on medical information was assessed using PROBAST (Table 2 and Figure 3). According to the PROBAST evaluation, only 2 studies [16,25] were designed as prospective cohort studies and were thus rated as low-ROB. In contrast, the majority of studies [13-15,17-24] were retrospective cohort studies and were consequently rated as high-risk. In the domains of predictors and outcomes, item 3.2 refers to whether the outcome was prespecified or defined using standard criteria. One study [15] lacked sufficient

information on this item and was therefore rated as having an “unclear” ROB. Additionally, 2 studies [16,20] failed to appropriately evaluate model performance; specifically, item 4.7, which assesses whether performance measures were adequately reported and validated, was rated as “high-risk” for these studies. Given that most of the included studies were retrospective in nature, with only a few being prospective, the overall ROB in outcome assessment was considered to be high or unclear across the reviewed literature.

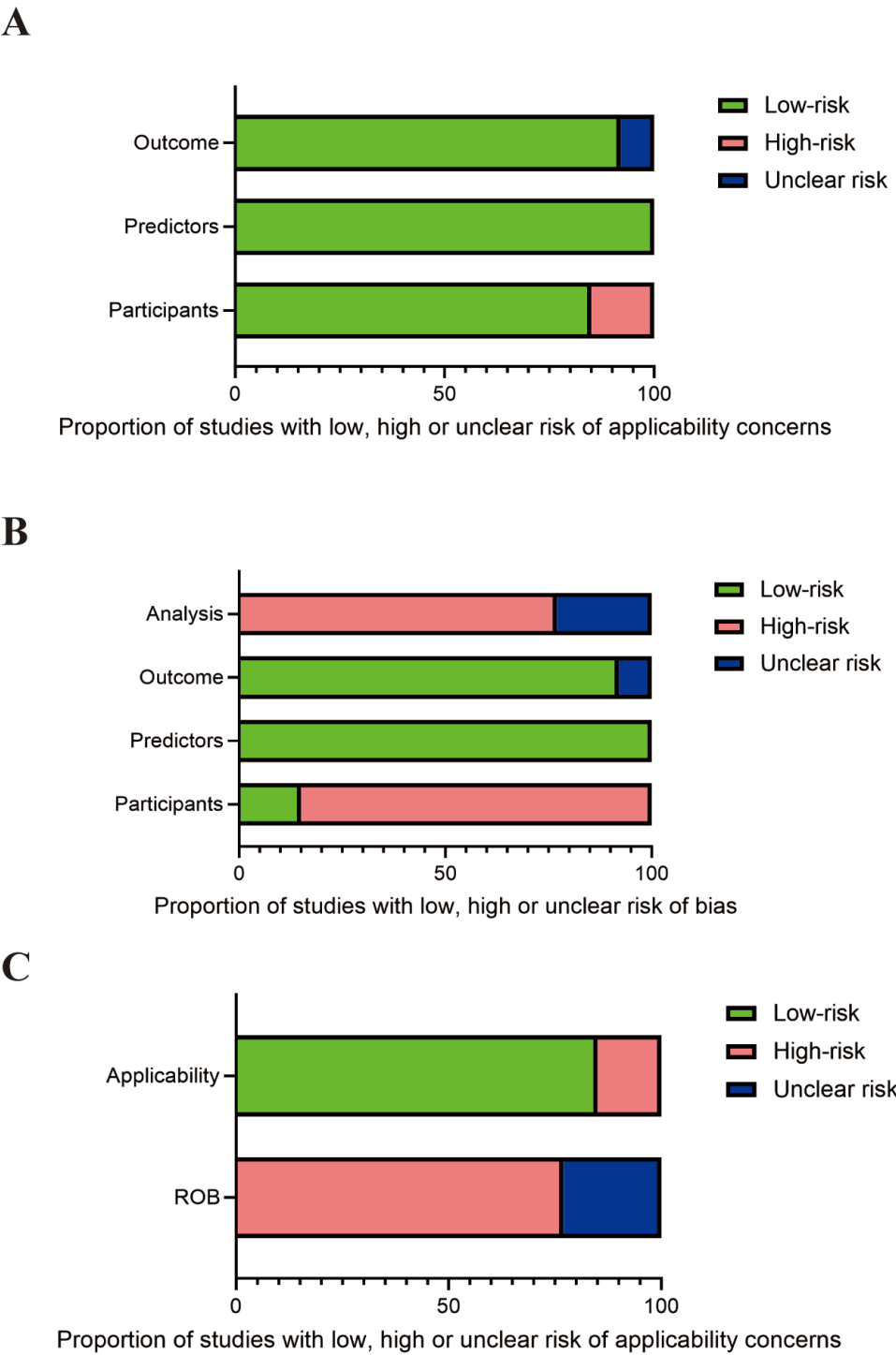
**Table 2.** Risk-of-bias assessment in machine learning prediction using Prediction Model Risk of Bias Assessment.

Study	Risk of bias				Applicability concerns			Overall	
	Participants	Predictors	Outcomes	Analysis	Participants	Predictors	Outcomes	ROB <sup>d</sup>	Applicability
Maki et al [13]	– <sup>a</sup>	+ <sup>b</sup>	+	–	+	+	+	–	+
Belliveau et al [14]	–	+	+	–	+	+	+	–	+
Kalyani et al [15]	–	+	? <sup>c</sup>	–	–	+	?	–	–
Facchinello et al [16]	+	+	+	?	+	+	+	?	+
Inoue et al [17]	–	+	+	–	+	+	+	–	+
Shimizu et al [18]	–	+	+	–	+	+	+	–	+
Okimatsu et al [19]	–	+	+	–	+	+	+	–	+
Luther et al [22]	–	+	+	–	–	+	+	–	–
Li et al [21]	–	+	+	–	+	+	+	–	+
Kato et al [20]	–	+	+	–	+	+	+	–	+

Study	Risk of bias				Applicability concerns			Overall	
Kim et al [23]	–	+	+	?	+	+	+	?	+
Hori et al [24]	–	+	+	?	+	+	+	?	+
DeVries et al [25]	+	+	+	–	+	+	+	–	+

<sup>a</sup>–: high-risk.  
<sup>b</sup>+: low-risk.  
<sup>c</sup>?: unclear risk.  
<sup>d</sup>ROB: risk of bias.

**Figure 3.** Schematic diagram of bias risk assessment components. (A) Proportion of studies with low, high, or unclear risk of bias; (B) proportion of studies with low, high, or unclear risk of applicability concerns; (C) proportion of studies with low, high, or unclear risk of applicability concerns. ROB: risk of bias.





Meta-Analysis Results

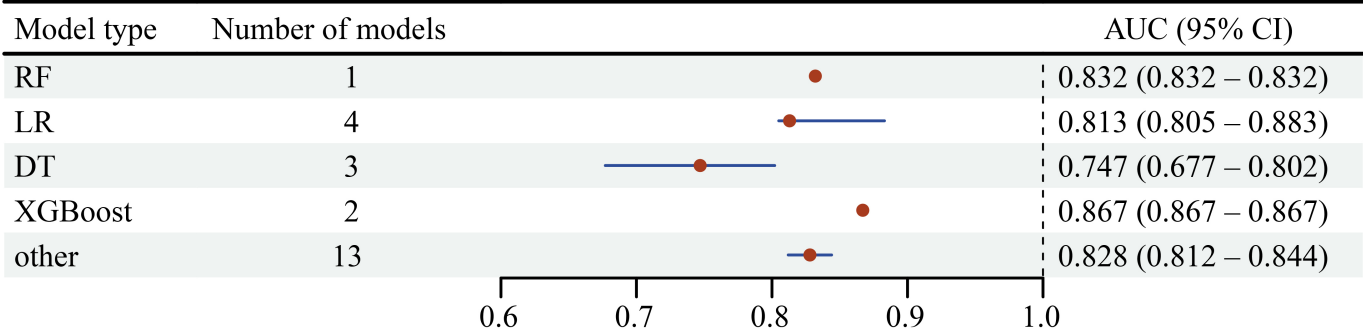
We conducted separate meta-analyses on different prognostic indicators. Overall, ML applications in SCI prognosis showed a high AUC. Considering that many studies lacked assessments of accuracy, sensitivity, and specificity, and that there were certain differences between algorithms and models, this reflects the complexity and significance of SCI disease prognosis prediction. Therefore, we only focused on and explored the AUC. The specific analysis results are listed below.

Spinal Cord Functional Prognosis

We included a total of 8 studies [13-20] with 7326 patients, using 17 ML algorithms to predict functional recovery

indicators, such as American Spinal Injury Association grading, motor scores, and sensory scores. Three studies [13, 19,20] used the random forest (RF) algorithm, 4 studies [13-15,17] used the logistic regression (LR) algorithm, and 3 studies [13,15,17] used the decision tree (DT) algorithm. The sample size for studies using the RF algorithm was 3547, for the LR algorithm was 6594, and for the DT algorithm was 3452. The AUC for the RF algorithm was 0.832, for the LR algorithm was 0.813 (95% CI 0.805-0.883), for the DT algorithm was 0.747 (95% CI 0.677-0.802), and for the extreme gradient boosting (XGBoost) algorithm was 0.867 (Figure 4).

Figure 4. Forest plot of models related to spinal cord functional prognosis. DT: decision tree; LR: logistic regression; RF: random forest; XGBoost: extreme gradient boosting.

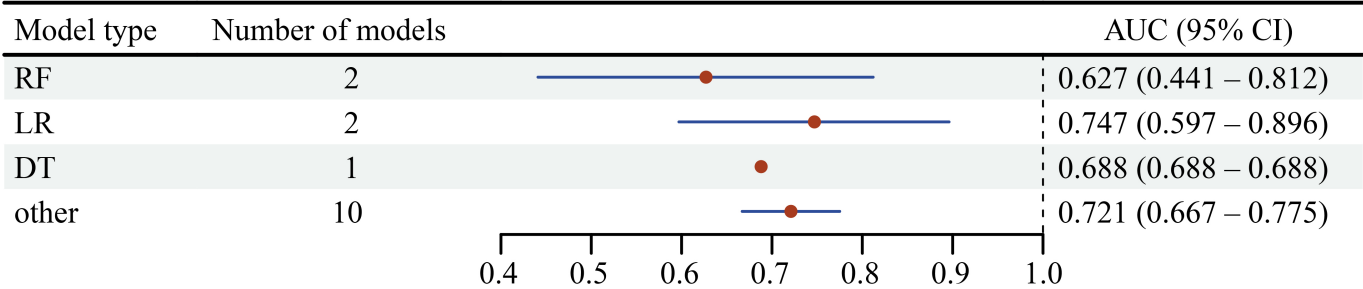


Postoperative Complications

We included 3 studies with a total of 6202 patients, using 12 ML algorithms. Two studies used RF and LR algorithms, with the RF algorithm having a sample size of 1493 and the

LR algorithm having a sample size of 5332. The AUC for the RF algorithm was 0.627 (95% CI 0.441-0.812), for the LR algorithm was 0.747 (95% CI 0.597-0.896), and for the DT algorithm was 0.688 (Figure 5).

Figure 5. Forest plot of models related to postoperative complications. RF: random forest; LR: logistic regression; DT: decision tree.



Independent Living Ability

A study [24] was conducted to assess the ability of patients with SCI to live independently. A total of 1404 patients were included, and the classification and regression tree (CART) algorithm was used. The model’s predictive variables included age, sex, Functional Independence Measure score, American Spinal Injury Association score, and the presence or absence of complications, such as diabetes and hypertension. The AUC of the CART model was 0.813.

Walking Ability

Currently, only 1 study [25] has been collected to evaluate the independent walking ability of patients with SCI at discharge or at a 1-year follow-up. This study included 862 patients and used LR, vector machine (VM), and Hicks algorithms for prediction, with AUCs of 0.720, 0.780, and 0.760, respectively. The model based on the VM algorithm showed the highest effectiveness.

## Discussion

### Principal Findings

SCI remains one of the leading causes of death and disability worldwide [26]. The heterogeneity factors, such as variable disease course, high inconsistency, and complex pathophysiological processes are major reasons for poor prognosis [27]. Clinically, a key concern for patients and their families is the prognosis, which is often assessed using imaging techniques, such as head computed tomography and magnetic resonance imaging, requiring considerable experience from clinicians. If clinicians had an objective assessment tool that could quickly and accurately predict prognosis based on the patient's condition, it would provide greater confidence in prognosis judgments [28]. Therefore, the idea of using ML to facilitate objective and individualized prognosis predictions for patients with SCI has emerged. In recent years, the application of ML in SCI prognosis prediction has made continuous breakthroughs, and the number and quality of related articles have been increasing over time. To our knowledge, this is the first meta-analysis evaluating the performance of ML models in predicting SCI prognosis. We conducted a comprehensive literature search and systematically evaluated the performance of ML models in SCI prognosis. Ultimately, ML algorithms were used to predict SCI prognosis in 13 included studies [13-25]. Among various model algorithms, although LR is a traditional algorithm, it is still commonly used in SCI prognosis prediction [13-15,17,22,23,25]. Other major algorithms include RF, DT, XGBoost, ANN, etc. A total of 18 algorithms were used to predict functional recovery after SCI, 12 algorithms for predicting complications after SCI, 1 algorithm for predicting independent living ability after SCI, and 3 algorithms for predicting walking ability after SCI. These models demonstrated varying levels of AUC. AUC indicates the algorithm's ability to balance sensitivity and specificity to minimize false negatives and false positives [29]. Therefore, its evaluation is particularly important in SCI prognosis assessment models. For predicting functional recovery, the ANN algorithm showed the highest AUC of 0.902. For predicting complications, the support vector machine-linear algorithm showed the highest AUC of 0.904. For predicting post-injury walking ability, both LR and VM models demonstrated the highest AUC of 0.870. Our study also found that over 90% of the included studies were published in the past 3 years, indicating growing interest in using ML-related predictive models to guide clinical decision-making.

It is worth noting that although most of the data in this study come from hospitals, the participant populations are from different countries and ethnic groups, with a focus on Asian and European populations, and there is little research on African populations. Published studies have not yet used predictive models to examine different regional and ethnic groups. Therefore, further model construction and validation using data from different populations are needed to increase confidence. Research indicates that ML models are significantly superior to traditional statistical models in handling large sample data [30,31]. However, most of the 13 studies

[13-25] included had sample sizes of fewer than 1000 cases, which suggests that future research should aim to collect sufficiently large and comprehensive sample sizes to reduce bias. Most studies opted to randomly split single-center data into training and testing sets or used different time periods for internal validation. However, the lack of external validation poses significant challenges to the model's applicability, potentially leading to overestimation of the model's effectiveness. Thus, SCI-related prognostic models require rigorous external validation to enhance their credibility.

Our study also indicates that most of the included studies use functional recovery as the outcome measure for SCI prognosis [13-20], and over time, research using ML has progressively reached higher quality standards. This has aroused considerable curiosity among SCI researchers and reflects the growing progress and recognition of ML applications in the field of SCI. In clinical practice, the AIS score remains of significant importance for evaluating the prognosis of patients with SCI. The AIS score uses clear criteria to assess patients' neurological function, offering objectivity and reproducibility and is applicable to various types of SCI, including traumatic and nontraumatic [32]. Despite its practicality, the AIS score has limitations. First, it mainly focuses on physiological aspects, such as sensory and motor functions, while neglecting the impact of psychological and social factors, which may affect clinicians' prognosis assessments. Second, it may lack sensitivity and might not assess mild SCI, thereby affecting long-term prognosis. Additionally, the AIS score is not suitable for SCI in children younger than 4 years [33].

In contrast, ML continues to evolve, and the improvement of various SCI databases highlights the increasing potential of ML in predicting SCI prognosis. ML can handle large amounts of complex data and extract nonlinear relationships, offering promising solutions to unknown challenges [34]. Our study found that 5 studies [13-15,17,24] used public databases to build predictive models and achieved good results in model performance. Additionally, many studies not only used traditional models, such as RF and DT, but also used recently popular models such as XGBoost. For instance, the XGBoost model constructed by Inoue et al [17] achieved an AUC of 0.867, outperforming models, such as LR and RF, and holds promise for future clinical guidance. Traditional clinical prognostic tools often face subjective biases of researchers and limited variables, while ML incorporates as many variables as possible, including demographic data, imaging, and laboratory results. This broad coverage allows for accurate and detailed individualized predictions based on each patient's specific situation, greatly reducing ambiguity and generalization. Furthermore, ML models continuously learn and optimize to improve prediction accuracy and explore potential patterns in data, potentially matching or even exceeding human brain capabilities [35].

### Limitations

This systematic review and meta-analysis have several limitations. First, the data quality from retrospective studies is lower compared to prospective studies. Among the 13



studies [13-25] included, only 2 were prospective, while the rest were retrospective, resulting in overall lower quality of the included literature. According to PROBAST guidelines, future research should aim to use prospective studies to obtain data, ensuring reliable model performance. Second, we focused solely on AUC to evaluate the performance of predictive models, without a comprehensive assessment using sensitivity, specificity, and  $F_1$ -score metrics. This is due to incomplete data in the selected literature, so the meta-analysis results may be somewhat one-sided. Moreover, because most studies did not consistently report key metrics, such as sensitivity and specificity, we did not perform a formal heterogeneity analysis. This limitation may affect the generalizability of the pooled AUC results across different study settings. Our study also did not focus much on the specific segments of the spinal cord damaged in patients with SCI or perform subgroup analysis based on different injury segments. Studies have shown that the mortality rate for patients with cervical SCI in the acute phase ranges from 11.2% to 25% [36], with the risk of death being about 7 times higher in upper cervical SCI compared to lower cervical SCI [37]. This indicates that prognosis varies with different spinal cord segments, suggesting a need for future research to focus on this aspect.

We must also recognize that ML research is inherently susceptible to multiple sources of bias that extend beyond the scope of conventional study design. The included studies used a wide range of algorithms, from traditional LR models to more complex methods, such as XGBoost and neural networks, resulting in significant heterogeneity in both model development and performance evaluation. Moreover, bias

may arise from imbalanced data distributions, insufficient feature representation, and lack of transparency in dataset construction. These issues are particularly concerning given that most studies did not undergo external validation, which further increases the risk of model overfitting and undermines the reliability of their clinical applicability. Additionally, model interpretability is an important factor to consider. To date, no ML model has been truly used for predicting SCI patient prognosis in clinical practice. Higher interpretability of a model allows clinicians to better understand its predictive value and make clinical decisions that benefit patients [38]. Lundberg et al [39] used the Shapley additive explanations algorithm to interpret any ML model's outputs and quantify each variable's contribution using Shapley additive explanations values, thereby promoting clinical application. Future research is expected to continue addressing this challenging issue of model interpretability [40].

## Conclusions

Increasing research efforts are devoted to developing SCI prognosis prediction models, which can provide personalized prognosis assessments for patients and ultimately alleviate their burden. However, the application of ML in this field is still in its early stages. Existing studies lack some data and details on predictive model reporting, and the majority are retrospective, leading to limited practical application of these models. Although clinical translation of ML is still a long way off, this study demonstrates that ML under artificial intelligence is poised to bring about a significant revolution in the field of SCI.

## Acknowledgments

The authors would like to thank all the researchers and participants involved for their contributions.

## Funding

This work was partially funded by the Medical Key Specialty Project of PLA Joint Logistics Team (LQZD-SW), Fujian Provincial Science and Technology Innovation Joint Fund Project (2024Y9644), and the Science and Technology Innovation Platform Project of Fujian Province (2022Y2017).

## Authors' Contributions

LZ, QH, HZ, and LW: concept and design.

LZ, QH, LX, YC, and JW: data collection, analysis, or interpretation.

LZ, QH, HZ, and LW: manuscript drafting.

LW: critical revision of the manuscript and technical or material support.

LZ: other protocol review.

## Conflicts of Interest

None declared.

## Checklist 1

PRISMA checklist.

[[DOCX File \(Microsoft Word File\), 24 KB-Checklist 1](#)]

## References

1. GBD Spinal Cord Injuries Collaborators. Global, regional, and national burden of spinal cord injury, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Neurol.* Nov 2023;22(11):1026-1047. [doi: [10.1016/S1474-4422\(23\)00287-9](#)] [Medline: [37863591](#)]

2. Dietz N, Alkin V, Mettelle J, Boakye M, Drazin D. Machine learning in clinical diagnosis, prognostication, and management of acute traumatic spinal cord injury (SCI): a systematic review. *J Clin Orthop Trauma*. Dec 2022;35:102046. [doi: [10.1016/j.jcot.2022.102046](https://doi.org/10.1016/j.jcot.2022.102046)] [Medline: [36425281](https://pubmed.ncbi.nlm.nih.gov/36425281/)]
3. Håkansson S, Tuci M, Bolliger M, Curt A, Jutzeler CR, Brünink SC. Data-driven prediction of spinal cord injury recovery: an exploration of current status and future perspectives. *Exp Neurol*. Oct 2024;380:114913. [doi: [10.1016/j.expneurol.2024.114913](https://doi.org/10.1016/j.expneurol.2024.114913)] [Medline: [39097073](https://pubmed.ncbi.nlm.nih.gov/39097073/)]
4. Choy W, Kyritsis N, Fernandez XD, et al. 188 American Spinal Injury Association (ASIA) Impairment Scale (AIS) conversion underestimates neurological recovery following traumatic spinal cord injury. *Neurosurgery*. 2023;69(Supplement\_1):29-29. [doi: [10.1227/neu.0000000000002375\\_188](https://doi.org/10.1227/neu.0000000000002375_188)]
5. Kirshblum S, Botticello A, Benedetto J, et al. A comparison of diagnostic stability of the ASIA Impairment Scale versus frankel classification systems for traumatic spinal cord injury. *Arch Phys Med Rehabil*. Sep 2020;101(9):1556-1562. [doi: [10.1016/j.apmr.2020.05.016](https://doi.org/10.1016/j.apmr.2020.05.016)] [Medline: [32531222](https://pubmed.ncbi.nlm.nih.gov/32531222/)]
6. Babalola TI, Yusuf SA, Raji M, Kamaldeen JO, Dolapo D. Value of the brain and spinal injury center score in assessment and prognosis of acute traumatic spinal cord injury. *Neurotrauma Rep*. 2024;5(1):592-605. [doi: [10.1089/neur.2023.0112](https://doi.org/10.1089/neur.2023.0112)] [Medline: [39071985](https://pubmed.ncbi.nlm.nih.gov/39071985/)]
7. Park EK, Lee KS, Seo BK, et al. Machine learning approaches to radiogenomics of breast cancer using low-dose perfusion computed tomography: predicting prognostic biomarkers and molecular subtypes. *Sci Rep*. Nov 28, 2019;9(1):17847. [doi: [10.1038/s41598-019-54371-z](https://doi.org/10.1038/s41598-019-54371-z)] [Medline: [31780739](https://pubmed.ncbi.nlm.nih.gov/31780739/)]
8. Gupta RK, Bharti S, Kunhare N, Sahu Y, Pathik N. Brain tumor detection and classification using cycle generative adversarial networks. *Interdiscip Sci*. Jun 2022;14(2):485-502. [doi: [10.1007/s12539-022-00502-6](https://doi.org/10.1007/s12539-022-00502-6)] [Medline: [35137330](https://pubmed.ncbi.nlm.nih.gov/35137330/)]
9. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature New Biol*. Aug 2021;596(7873):583-589. [doi: [10.1038/s41586-021-03819-2](https://doi.org/10.1038/s41586-021-03819-2)] [Medline: [34265844](https://pubmed.ncbi.nlm.nih.gov/34265844/)]
10. Hanna MG, Pantanowitz L, Jackson B, et al. Ethical and bias considerations in artificial intelligence/machine learning. *Mod Pathol*. Mar 2025;38(3):100686. [doi: [10.1016/j.modpat.2024.100686](https://doi.org/10.1016/j.modpat.2024.100686)] [Medline: [39694331](https://pubmed.ncbi.nlm.nih.gov/39694331/)]
11. Fernandez-Felix BM, López-Alcalde J, Roqué M, Muriel A, Zamora J. CHARMS and PROBAST at your fingertips: a template for data extraction and risk of bias assessment in systematic reviews of predictive models. *BMC Med Res Methodol*. Feb 17, 2023;23(1):44. [doi: [10.1186/s12874-023-01849-0](https://doi.org/10.1186/s12874-023-01849-0)] [Medline: [36800933](https://pubmed.ncbi.nlm.nih.gov/36800933/)]
12. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med*. Jan 1, 2019;170(1):W1-W33. [doi: [10.7326/M18-1377](https://doi.org/10.7326/M18-1377)] [Medline: [30596876](https://pubmed.ncbi.nlm.nih.gov/30596876/)]
13. Maki S, Furuya T, Inoue T, et al. Machine learning web application for predicting functional outcomes in patients with traumatic spinal cord injury following inpatient rehabilitation. *J Neurotrauma*. May 2024;41(9-10):1089-1100. [doi: [10.1089/neu.2022.0383](https://doi.org/10.1089/neu.2022.0383)] [Medline: [37917112](https://pubmed.ncbi.nlm.nih.gov/37917112/)]
14. Belliveau T, Jette AM, Seetharama S, et al. Developing artificial neural network models to predict functioning one year after traumatic spinal cord injury. *Arch Phys Med Rehabil*. Oct 2016;97(10):1663-1668. [doi: [10.1016/j.apmr.2016.04.014](https://doi.org/10.1016/j.apmr.2016.04.014)] [Medline: [27208647](https://pubmed.ncbi.nlm.nih.gov/27208647/)]
15. Kalyani P, Manasa Y, Ahammad SH, et al. Prediction of patient's neurological recovery from cervical spinal cord injury through XGBoost learning approach. *Eur Spine J*. Jun 2023;32(6):2140-2148. [doi: [10.1007/s00586-023-07712-6](https://doi.org/10.1007/s00586-023-07712-6)] [Medline: [37060466](https://pubmed.ncbi.nlm.nih.gov/37060466/)]
16. Facchinello Y, Beauséjour M, Richard-Denis A, Thompson C, Mac-Thiong JM. Use of regression tree analysis for predicting the functional outcome after traumatic spinal cord injury. *J Neurotrauma*. May 1, 2021;38(9):1285-1291. [doi: [10.1089/neu.2017.5321](https://doi.org/10.1089/neu.2017.5321)] [Medline: [29065782](https://pubmed.ncbi.nlm.nih.gov/29065782/)]
17. Inoue T, Ichikawa D, Ueno T, et al. XGBoost, a machine learning method, predicts neurological recovery in patients with cervical spinal cord injury. *Neurotrauma Rep*. 2020;1(1):8-16. [doi: [10.1089/neur.2020.0009](https://doi.org/10.1089/neur.2020.0009)] [Medline: [34223526](https://pubmed.ncbi.nlm.nih.gov/34223526/)]
18. Shimizu T, Suda K, Maki S, et al. Efficacy of a machine learning-based approach in predicting neurological prognosis of cervical spinal cord injury patients following urgent surgery within 24 h after injury. *J Clin Neurosci*. Jan 2023;107:150-156. [doi: [10.1016/j.jocn.2022.11.003](https://doi.org/10.1016/j.jocn.2022.11.003)] [Medline: [36376152](https://pubmed.ncbi.nlm.nih.gov/36376152/)]
19. Okimatsu S, Maki S, Furuya T, et al. Determining the short-term neurological prognosis for acute cervical spinal cord injury using machine learning. *J Clin Neurosci*. Feb 2022;96:74-79. [doi: [10.1016/j.jocn.2021.11.037](https://doi.org/10.1016/j.jocn.2021.11.037)] [Medline: [34998207](https://pubmed.ncbi.nlm.nih.gov/34998207/)]
20. Kato C, Uemura O, Sato Y, Tsuji T. Functional outcome prediction after spinal cord injury using ensemble machine learning. *Arch Phys Med Rehabil*. Jan 2024;105(1):95-100. [doi: [10.1016/j.apmr.2023.08.011](https://doi.org/10.1016/j.apmr.2023.08.011)] [Medline: [37714506](https://pubmed.ncbi.nlm.nih.gov/37714506/)]
21. Li MP, Liu WC, Wu JB, et al. Machine learning for the prediction of postoperative nosocomial pulmonary infection in patients with spinal cord injury. *Eur Spine J*. Nov 2023;32(11):3825-3835. [doi: [10.1007/s00586-023-07772-8](https://doi.org/10.1007/s00586-023-07772-8)] [Medline: [37195363](https://pubmed.ncbi.nlm.nih.gov/37195363/)]

22. Luther SL, Thomason SS, Sabharwal S, et al. Machine learning to develop a predictive model of pressure injury in persons with spinal cord injury. *Spinal Cord*. Sep 2023;61(9):513-520. [doi: [10.1038/s41393-023-00924-z](https://doi.org/10.1038/s41393-023-00924-z)] [Medline: [37598263](https://pubmed.ncbi.nlm.nih.gov/37598263/)]
23. Kim Y, Lim M, Kim SY, et al. Integrated machine learning approach for the early prediction of pressure ulcers in spinal cord injury patients. *J Clin Med*. Feb 8, 2024;13(4):990. [doi: [10.3390/jcm13040990](https://doi.org/10.3390/jcm13040990)] [Medline: [38398304](https://pubmed.ncbi.nlm.nih.gov/38398304/)]
24. Hori T, Imura T, Tanaka R. Development of a clinical prediction rule for patients with cervical spinal cord injury who have difficulty in obtaining independent living. *Spine J*. Feb 2022;22(2):321-328. [doi: [10.1016/j.spinee.2021.08.010](https://doi.org/10.1016/j.spinee.2021.08.010)] [Medline: [34487911](https://pubmed.ncbi.nlm.nih.gov/34487911/)]
25. DeVries Z, Hoda M, Rivers CS, et al. Development of an unsupervised machine learning algorithm for the prognostication of walking ability in spinal cord injury patients. *Spine J*. Feb 2020;20(2):213-224. [doi: [10.1016/j.spinee.2019.09.007](https://doi.org/10.1016/j.spinee.2019.09.007)] [Medline: [31525468](https://pubmed.ncbi.nlm.nih.gov/31525468/)]
26. Eli I, Lerner DP, Ghogawala Z. Acute traumatic spinal cord injury. *Neurol Clin*. May 2021;39(2):471-488. [doi: [10.1016/j.necl.2021.02.004](https://doi.org/10.1016/j.necl.2021.02.004)] [Medline: [33896529](https://pubmed.ncbi.nlm.nih.gov/33896529/)]
27. Anjum A, Yazid MD, Fauzi Daud M, et al. Spinal cord injury: pathophysiology, multimolecular interactions, and underlying recovery mechanisms. *Int J Mol Sci*. Oct 13, 2020;21(20):7533. [doi: [10.3390/ijms21207533](https://doi.org/10.3390/ijms21207533)] [Medline: [33066029](https://pubmed.ncbi.nlm.nih.gov/33066029/)]
28. Yang Y, Tang L, Deng Y, et al. The predictive performance of artificial intelligence on the outcome of stroke: a systematic review and meta-analysis. *Front Neurosci*. 2023;17:1256592. [doi: [10.3389/fnins.2023.1256592](https://doi.org/10.3389/fnins.2023.1256592)] [Medline: [37746141](https://pubmed.ncbi.nlm.nih.gov/37746141/)]
29. Moulaei K, Sharifi H, Bahaadinbeigy K, Haghdoust AA, Nasiri N. Machine learning for prediction of viral hepatitis: a systematic review and meta-analysis. *Int J Med Inform*. Nov 2023;179:105243. [doi: [10.1016/j.ijmedinf.2023.105243](https://doi.org/10.1016/j.ijmedinf.2023.105243)] [Medline: [37806178](https://pubmed.ncbi.nlm.nih.gov/37806178/)]
30. Kaplan A, Cao H, FitzGerald JM, et al. Artificial intelligence/machine learning in respiratory medicine and potential role in asthma and COPD diagnosis. *J Allergy Clin Immunol Pract*. Jun 2021;9(6):2255-2261. [doi: [10.1016/j.jaip.2021.02.014](https://doi.org/10.1016/j.jaip.2021.02.014)] [Medline: [33618053](https://pubmed.ncbi.nlm.nih.gov/33618053/)]
31. Jammal AA, Thompson AC, Mariottoni EB, et al. Human versus machine: comparing a deep learning algorithm to human gradings for detecting glaucoma on fundus photographs. *Am J Ophthalmol*. Mar 2020;211:123-131. [doi: [10.1016/j.ajo.2019.11.006](https://doi.org/10.1016/j.ajo.2019.11.006)] [Medline: [31730838](https://pubmed.ncbi.nlm.nih.gov/31730838/)]
32. Furbish CL, Anderson RK, Field-Fote EC. Accuracy of self-reported severity and level of spinal cord injury. *Spinal Cord*. Oct 2022;60(10):934-941. [doi: [10.1038/s41393-022-00855-1](https://doi.org/10.1038/s41393-022-00855-1)] [Medline: [36097066](https://pubmed.ncbi.nlm.nih.gov/36097066/)]
33. Qi Q, Wang L, Yang B, et al. The relationship between the structural changes in the cervical spinal cord and sensorimotor function of children with thoracolumbar spinal cord injury (TLSCI). *Spinal Cord*. Jul 2024;62(7):414-420. [doi: [10.1038/s41393-024-01000-w](https://doi.org/10.1038/s41393-024-01000-w)] [Medline: [38824252](https://pubmed.ncbi.nlm.nih.gov/38824252/)]
34. An Q, Rahman S, Zhou J, Kang JJ. A comprehensive review on machine learning in healthcare industry: classification, restrictions, opportunities and challenges. *Sensors (Basel)*. Apr 22, 2023;23(9):4178. [doi: [10.3390/s23094178](https://doi.org/10.3390/s23094178)] [Medline: [37177382](https://pubmed.ncbi.nlm.nih.gov/37177382/)]
35. Binson VA, Thomas S, Subramoniam M, Arun J, Naveen S, Madhu S. A review of machine learning algorithms for biomedical applications. *Ann Biomed Eng*. May 2024;52(5):1159-1183. [doi: [10.1007/s10439-024-03459-3](https://doi.org/10.1007/s10439-024-03459-3)] [Medline: [38383870](https://pubmed.ncbi.nlm.nih.gov/38383870/)]
36. Jimenez-Mesa C, Arco JE, Martinez-Murcia FJ, Suckling J, Ramirez J, Gorris JM. Applications of machine learning and deep learning in SPECT and PET imaging: general overview, challenges and future prospects. *Pharmacol Res*. Nov 2023;197:106984. [doi: [10.1016/j.phrs.2023.106984](https://doi.org/10.1016/j.phrs.2023.106984)] [Medline: [37940064](https://pubmed.ncbi.nlm.nih.gov/37940064/)]
37. Sadeghi-Naini M, Yousefifard M, Ghodsi Z, et al. In-hospital mortality rate in subaxial cervical spinal cord injury patients: a systematic review and meta-analysis. *Acta Neurochir (Wien)*. Sep 2023;165(9):2675-2688. [doi: [10.1007/s00701-023-05720-5](https://doi.org/10.1007/s00701-023-05720-5)] [Medline: [37480505](https://pubmed.ncbi.nlm.nih.gov/37480505/)]
38. Fehlings MG, Tetreault LA, Wilson JR, et al. A clinical practice guideline for the management of patients with acute spinal cord injury and central cord syndrome: recommendations on the timing ( $\leq 24$  hours versus  $>24$  hours) of decompressive surgery. *Global Spine J*. Sep 2017;7(3 Suppl):195S-202S. [doi: [10.1177/2192568217706367](https://doi.org/10.1177/2192568217706367)] [Medline: [29164024](https://pubmed.ncbi.nlm.nih.gov/29164024/)]
39. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. Jan 2020;2(1):56-67. [doi: [10.1038/s42256-019-0138-9](https://doi.org/10.1038/s42256-019-0138-9)] [Medline: [32607472](https://pubmed.ncbi.nlm.nih.gov/32607472/)]
40. Magrabi F, Ammenwerth E, McNair JB, et al. Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. *Yearb Med Inform*. Aug 2019;28(1):128-134. [doi: [10.1055/s-0039-1677903](https://doi.org/10.1055/s-0039-1677903)] [Medline: [31022752](https://pubmed.ncbi.nlm.nih.gov/31022752/)]

## Abbreviations

**AIS:** Abbreviated Injury Scale

**AUC:** area under the curve

**CART:** classification and regression tree

**CHARMS:** Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies

**DT:** decision tree

**LR:** logistic regression

**ML:** machine learning

**PRISMA-DTA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Diagnostic Test Accuracy

**PROBAST:** Prediction Model Risk of Bias Assessment Tool

**RF:** random forest

**ROB:** risk of bias

**SCI:** spinal cord injury

**VM:** vector machine

**XGBoost:** extreme gradient boosting

*Edited by Khaled El Emam; peer-reviewed by Carsten Weerth, Fred Geisler, Kwang-Sig Lee, Vedamurthy Gejjegondanahalli Yogeshappa; submitted 06.Sep.2024; final revised version received 16.Aug.2025; accepted 17.Aug.2025; published 05.Dec.2025*

*Please cite as:*

Zhong L, Huang Q, Zhang H, Xue L, Chen Y, Wu J, Wei L

Predicting Spinal Cord Injury Prognosis Using Machine Learning: Systematic Review and Meta-Analysis

JMIR AI2025;4:e66233

URL: <https://ai.jmir.org/2025/1/e66233>

doi: [10.2196/66233](https://doi.org/10.2196/66233)

© Linxing Zhong, Qiying Huang, Hao Zhang, Liang Xue, Yehuang Chen, Jianwu Wu, Liangfeng Wei. Originally published in JMIR AI (<https://ai.jmir.org>), 05.Dec.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR AI, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.ai.jmir.org/>, as well as this copyright and license information must be included.